ADHDgene: a genetic database for attention deficit hyperactivity disorder

Liuyan Zhang[1,2], Suhua Chang[1,2], Zhao Li[1,2], Kunlin Zhang[1], Yang Du[1,2], Jurg Ott[1] and Jing Wang[1,*]

[1]Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101 and [2]Graduate University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

With a worldwide prevalence of ∼5%, attention deficit hyperactivity disorder (ADHD) has become one of the most common psychiatric disorders. The polygenetic nature of ADHD indicates that multiple genes jointly contribute to the development of this complex disease. Studies aiming to explore genetic susceptibility of ADHD have been increasing in recent years. There is a growing need to integrate the genetic data from various genetic studies to provide a comprehensive data set and uniform access for convenience of in-depth data mining. So far, there has been no such effort for ADHD. To address the genetic complexity of ADHD, we developed the ADHDgene database by integrating ADHD-related genetic factors by profound literature reading. Based on the data from the literature, extended functional analysis, including linkage disequilibrium analysis, pathway-based analysis and gene mapping were performed to provide new insights into genetic causes of ADHD. Moreover, powerful search tools and a graphical browser were developed to facilitate the navigation of the data and data connections. As the first genetic database for ADHD, ADHDgene aims to provide researchers with a central genetic resource and analysis platform for ADHD and is freely available at http://adhd.psych.ac.cn/.

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is one of the most common psychiatric disorders with a worldwide prevalence of ∼5% (1). The prevalence is even higher among school-age children ranging from 8% to 12% (2,3).

ADHD is predominantly childhood-onset and can persist into adolescence and adulthood to inflict long-term harm. It is characterized by a continuous and combined pattern of inattentive, hyperactive and impulsive behavior, and is often comorbid with other psychiatric disorders (4). Patients with ADHD have impaired academic, executive and social functions, which also lead to a serious financial burden to families and society. Although the etiology of ADHD is still incompletely understood, results from family, twin and adoption studies, as well as molecular genetic studies consistently indicate the strong genetic influence on ADHD with estimated heritability ranging from 75% to 91% (5). To date, the polygenetic nature of ADHD is widely acknowledged, which indicates multiple genes of moderate effect are involved in the genetic basis of ADHD (6). Meanwhile, alternative modes of inheritance in ADHD remain a topic for further clarification. For example, evidences from the segregation analysis demonstrated that a few major genes convey susceptibility to ADHD and they exhibit Mendelian segregation (7). Thus, to study the genetic basis of ADHD is of fundamental importance in uncovering disease mechanism and in developing effective methods for ADHD diagnosis, treatment and prevention.

During the past years, numerous studies aiming to explore genetic susceptibility of ADHD have been published. Large numbers of association studies (8), linkage studies (9) and meta-analyses (10) have been conducted, and numbers of susceptibility variants, genes and chromosomal regions have been reported to be associated with ADHD. For example, both genome-wide linkage and fine mapping studies support the linkage between ADHD and chromosome bands 16p13 (11,12). Genes of the dopaminergic and serotonergic system, including *DRD4*, *SLC6A3* (*DAT1*) and *DBH*, were widely studied and implicated to be associated with the susceptibility of ADHD. Several genome-wide association studies (GWAS) have been conducted (13), and studies of

---

genomic DNA copy number variants (CNVs) in ADHD have also emerged in recent years and identified some rare or large deletions/duplications in ADHD patients (14–16). However, these results are scattered in numerous publications and sometimes are equivocal or inconsistent. For example, a candidate–gene association study using a sample of 674 families indicated that both genes, *DRD4* and *SLC6A3*, are associated with ADHD respectively (17), while results from a case–control study did not find any association between these two genes and ADHD (18). As these inconsistent results are common in genetic studies, it would be difficult for researchers to acquire a global understanding of these positive and negative findings, and then design the right kind of study to move forward. The difficulty will be significantly increased with the fast growth of genetic data. Thus, to address the genetic complexity of ADHD and the heterogeneity of studies, a comprehensive and well-organized collection of genetic data from multiple published studies is urgently needed. Moreover, an association study, as the currently most favored and widely used approach for genetic study, usually identifies trait-associated genetic markers rather than causal variants affecting the trait. Regular data analysis of an association study examines SNPs/genes independently and ignores the combined SNP/gene effects. In order to take the best advantage of the current genetic data, in-depth data mining is urgently needed to provide new candidates and new insights into the mechanism of ADHD.

Here, we present ADHDgene, a genetic database for ADHD, to fulfill the growing needs of data integration and data analysis. Currently, there are several genetic databases focusing on psychiatric disorders, like the database of genetic studies of bipolar disorder (19), SzGene (20) and SZGR (21) for schizophrenia, and AutDB (22) and AGD (23) for autism. However, there has been no such effort for ADHD. As the first genetic database for ADHD, ADHDgene aims to provide the research community with a comprehensive set of genetic data from both profound literature screening and extended functional analysis. ADHDgene also provides online tools to search and browse different data types and data connections. In general, ADHDgene is designed to be a central genetic database for ADHD to facilitate the study of function and mechanism to help unveil the genetic basis of ADHD.

DATA CONTENT, DATA INTEGRATION AND ANALYSIS

Data content

The key data types in ADHDgene include variants (SNP, CNV, etc.), genes and chromosomal regions. There are mainly two sets of data hosted by ADHDgene. One is the *core data*, which was derived from full-text literature reading with manual curation of ADHD genetic studies from the PubMed database at NCBI (http://www.ncbi.nih.gov/pubmed). The other is the *extended data* derived from extended functional analysis based on the *core data*, including linkage disequilibrium (LD) analysis of the

literature-origin SNPs, pathway-based analysis (PBA) for the GWAS of ADHD and gene mapping. A full annotation was made for both data sets, including SNP functional annotation (such as non-synonymous coding SNPs, or SNPs leading to gain or lost of stop codon) and GO and KEGG pathway annotation for genes. Public databases utilized for the mapping and annotation include dbSNP (24), Ensembl (25), HGNC (26), UCSC (27), GO (28) and KEGG (29). The data statistics of ADHDgene dated 1 August 2011 is shown in Table 1. All data can be freely downloaded from our website.

Core data: integration and curation

The *core data* of ADHDgene integrates ADHD-related genetic factors from published literature. Besides SNP and CNV, other variants like variable number tandem repeat (VNTR) and microsatellite are also included. To obtain the *core data*, a comprehensive search of ADHD related genetic publications in PubMed was made by using the following search term: ('attention deficit hyperactivity disorder' [Title/Abstract] OR ADHD [Title/Abstract]) AND (gene [Title/Abstract] OR genetic [Title/Abstract] OR genome [Title/Abstract] OR genomic [Title/Abstract]) AND (linkage [Title/Abstract] OR association [Title/Abstract] OR polymorphism [Title/Abstract] OR SNP [Title/Abstract] OR variant [Title/Abstract] OR variation [Title/Abstract] OR mutation [Title/Abstract]). It resulted in ∼780 English publications from the year 2000 through 30 June 2011. By manual screening of abstracts of these publications, only publications about association studies, linkage studies and meta-analyses were kept for further reading. Other publications like reviews or articles about pharmacology, sociology, electrophysiology, neurophysiology and behavioral research were excluded. After filtering, there are 288 articles included in ADHDgene. The full text of each eligible publication was read carefully, and detailed information of each study was extracted manually. To ensure a comprehensive review, both positive and negative results were collected.

To illustrate the association between genetic candidates and ADHD, statistical results were further evaluated and were categorized into 'Significant', 'Non-significant' and 'Trend' according to their statistical evidence in the original publications. For linkage study, the significance levels were assigned based on the criteria proposed by Lander and Kruglyak (30). For candidate–gene association studies, a significance level of $P < 0.05$ was used unless the authors suggested some other value. The result of statistical analysis is defined as 'Significant' if the corresponding P-value is <0.05. For GWAS, $P < 1 \times 10^{-8}$ indicates a 'Significant' result, $P > 1 \times 10^{-5}$ indicates a 'Non-significant' result, and a value between these boundaries represents a 'Trend' result. Mutations identified in ADHD patients were all classified as 'Trend' unless statistical significance was proposed. Meanwhile, to help understand the results from statistical analysis, for each study, the study design, sample population, analytical method, as well as brief comments from authors, were presented to provide researchers a clear picture for each genetic factor.

Table 1. ADHDgene data content and statistics as of 1 August 2011

Data set[a]	Data type		Data statistics[b]
Core data	Variant	SNP	941 (291)
		CNV	296 (NA)
		Others[c]	160 (41)
	Gene		213 (83)
	Region		127 (29)
Extended data	Variant	LD-proxy	4120
	Gene	Mapped by published SNP	65
		Mapped by LD-proxy	140
		Mapped by CNV	891
		Mapped by significant region	2295
		In pathway identified by PBA	260
	Pathway	Pathway identified by PBA	8

[a]The *core data* is entirely from the literature, while the *extended data* is from extended functional analysis
[b]Number in parenthesis indicates the number of statistically significant results
[c]Other variants include VNTR, microsatellite, STR, duplication, SNP without reference SNP ID in dbSNP, etc. CNV, copy number variation; PBA, pathway-based analysis; NA, not applicable.

Extended data: extended functional analysis

The polygenetic nature of ADHD indicates that multiple genes jointly contribute to the development of ADHD with each having a modest effect on the overall risk (31). Although current research has reported many candidate SNPs and genes for ADHD, few are conclusive due to the low replication rate, or lack of functional evidence on mechanism through which the susceptibility genes jointly affect the trait. With the purpose to identify more candidate SNPs and genes, as well as to allow a deeper interpretation of current data, we performed extended functional analysis based on the *core data*, which includes LD analysis, pathway-based analysis and gene mapping. These analyses enriched the content of the database to provide new clues for understanding the pathological mechanism of ADHD.

Linkage disequilibrium (LD) analysis. Association studies have identified many SNPs as candidate genetic markers for ADHD, but most of them lack function, e.g. some intronic SNPs, flanking SNPs or synonymous coding SNPs. Linkage disequilibrium (LD) analysis searches the LD-proxies of the literature-origin SNPs with the purpose to capture more candidate SNPs, especially those candidate causal SNPs based on functional annotation. These extended LD-proxies will provide new candidate SNPs for future study designs.

The LD data used in this analysis were downloaded from the HapMap website, which was compiled from merged genotype data from phases I+II+III (HapMap rel #27, NCBI B36) for markers up to 200 kb apart (32). SNPs in LD ($r^2 > 0.8$) with published SNPs were defined as LD-proxies. The population used in the LD analysis is consistent with original studies. For example, the LD-proxies of SNP rs2873804, which was identified in the Chinese population, were extracted from HapMap LD data in the CHB, CHD and JPT populations (32).

Pathway-based analysis (PBA) for genome-wide association study (GWAS). So far, five genome-wide

association studies (GWASs) have been conducted for ADHD but few significant SNPs were found (33–37). As the original GWAS analysis focuses on single SNP/gene, combined SNP/gene effects are hardly explored, which leads to difficulties in understanding the biological function and mechanism of complex diseases (38). To overcome this limitation, we implemented a pathway-based analysis (PBA) algorithm, named *i*-GSEA (improved gene set enrichment analysis) developed in our recent studies (39,40), on the full list of GWAS SNP *P*-values to detect pathways/gene sets associated with ADHD. *i*-GSEA utilizes a re-scaled enrichment score (*ES*) to emphasize pathways with well evaluated high proportion of significant genes to identify pathways/gene sets with combined effects of modest genes involved in complex disease. False discovery rate (FDR) is calculated for each pathway/gene set and pathways/gene sets with FDR < 0.05 are regarded as statistical significance.

The ADHD GWAS data sets (full list of SNP *P*-values) were requested by emailing the corresponding authors of related publications or through dbGaP (41). Finally, two GWAS data sets were authorized. One is from a family-based GWAS using SNP array (35), and the other is from a case–control GWAS with pooled DNA (34). *i*-GSEA was performed on the above two data sets by using default parameters (39) with annotated GO terms from MSigDB 3.0 (42) as reference database. In total, the PBA analyses resulted in 8 pathways/gene sets significantly associated with ADHD (FDR < 0.05). Among them, the pathway annotated by GO term GO:0005262 represents a set of genes related to calcium channel activity, which has been reported to infer the etiology of other psychiatric disorders (43,44). This result provides new evidence for possible shared biological mechanisms between ADHD and other psychiatric disorders.

Gene mapping. To ensure a comprehensive collection of candidate genes for ADHD, published SNPs, which were not mapped to genes in the original publications, and LD-proxies from LD analysis were mapped to genes according to their chromosomal locations. Mapping

between genes and the reported CNVs and significant chromosomal regions were performed as well. All genes acquired by the above mappings, as well as genes in pathways identified by PBA, were regarded as the extended candidate genes for ADHD. This part of genes is distinguished as 'genes from other sources' in comparison to the 'literature-origin genes' from the *core data* in ADHDgene database.

DATABASE USAGE AND ACCESS

ADHDgene provides researchers with a powerful search engine and a user-friendly interface to access and browse different data types and data connections. To search the database efficiently, besides 'Quick Search' by keyword, ADHDgene offers 'Advanced Search' for variant, gene, region, pathway and publication to allow users to specify and combine query options. Particularly, to facilitate a thorough investigation on the *core data*, ADHDgene supports 'Cross Search' between different genetic factors (variant, gene and region) and publication, which helps to provide a thorough overview for each genetic factor. For example, users may want to know how many statistically significant associations have been reported for gene dopamine beta-hydroxylase (*DBH*) in the young generations of Caucasian populations. Through 'Cross Search', users may input the gene name '*DBH*', select 'Significant' in the field 'Result of statistical analysis', input 'Caucasian' in the field 'Population' and select 'Children/Adolescents' in the field 'Age group'.

ADHDgene provides a detailed report for each genetic factor. Taking the gene *DBH* for example, on its 'Gene Report', *DBH*-related 16 publications and key results from each publication are listed. Among them, nine articles reported significant associations with ADHD, and seven reported non-significant associations. To assess the heterogeneity of these studies, users may link to the 'Publication Report' to view the details. Other genetic factors related to *DBH* are all listed, such as SNPs (literature-origin or LD-proxies), CNVs, regions, and genes sharing same pathway with *DBH* (by database annotation or PBA, which indicate potential candidates for epistasis study). It is worth mentioning that there are 12 LD-proxies identified for *DBH*. Five of them have been reported to have linkage or association signals with other diseases, like schizophrenia (45,46) and Parkinson's disease (47). It is likely that these LD-proxies could be possible candidate markers for ADHD.

Besides the search module, GBrowse (48) was incorporated in ADHDgene to facilitate viewing the different types of genetic factors (variant/gene/region) simultaneously in the context of genomic regions. GBrowse is a popular visualization tool for visualizing genetic and genomic data. Entries from the literature and extended functional analysis are marked in different colors in the browser. Users can also link to the detail page of corresponding entries from the browser. In summary, through the above search and browse tools, users can access and view the logical connections between different genetic

factors with strong evidence support either from the literature or from functional analysis (Figure 1).
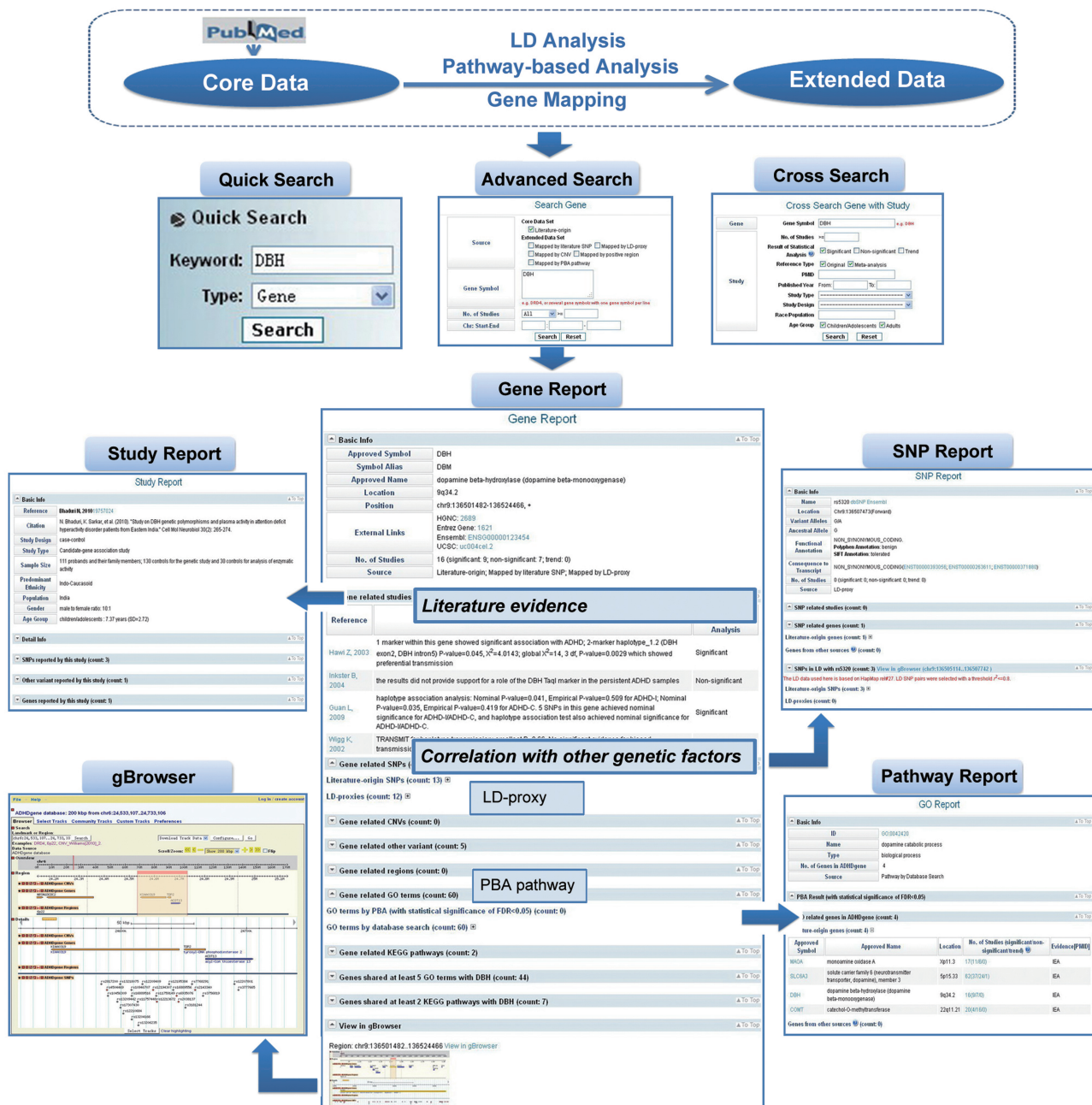
SYSTEM DESIGN AND IMPLEMENTATION

All data of ADHDgene are stored and managed in a MySQL relational database. The website is implemented using JSP running on an Apache Tomcat web server. Struts framework and Hibernate were employed to help improve the stability of the web services. AJAX and some jQuery plugins were used for the interface development. The data analysis programs were written by PERL. GBrowse uses MySQL as backend and was built following the configuration files provided by its developer (http://gmod.org/wiki/GBrowse_Configuration_HOWTO). The genomic data in ADHDgene are loaded into GBrowse after being converted into genome feature format (GFF).

DISCUSSION AND FUTURE DEVELOPMENT

As the first genetic database for ADHD, ADHDgene was developed to provide a panoramic view of present genetic studies for this disease. ADHDgene covers a broad range of data types including not only SNPs, genes, chromosomal regions and pathways, but also variants like CNV, VNTR and microsatellites. By integrating data from both publications and extended functional analysis, ADHDgene provides a comprehensive genetic data set for ADHD. In comparison to other disease-centered genetic databases, ADHDgene was developed in a reliable (manual curation of literature) but not limited (extended functional analysis) way to fulfill the increasing research demands in addressing the genetic complexity of ADHD. By providing powerful search and browse tools, the ADHDgene database aims to act as not only an integrated genetic resource for ADHD, but also as a flexible application platform for the genetic study of ADHD.

In the next few years, the number of genetic studies of ADHD is expected to keep increasing especially with the development of new technologies, such as next generation sequencing (49). ADHDgene will be periodically updated to ensure a most up-to-date follow up of the genetic research progress of ADHD. It should be noted that pathways by PBA in ADHDgene provide potential hypotheses for further study rather than conclusive results, since the results from PBA might be affected by the quality of GWAS data (50). As soon as additional high quality GWAS data become available, we will analyze them to update the PBA results in ADHDgene. Meanwhile, as studies of rare variants, epigenetics, large-scale gene expression and gene–environment interplay for ADHD are expected to accelerate, the scope of ADHDgene will be expanded to integrate newly generated data. Evidence from animal models will also be one of the future efforts in updating ADHDgene. We hope our continuous efforts will help to unveil the genetic basis of ADHD and to contribute to global mental health.

Figure 1. Screenshot of the search and browse tools for access and analysis of data and data connections in ADHDgene.

REFERENCES

1. Polanczyk,G., de Lima,M.S., Horta,B.L., Biederman,J. and Rohde,L.A. (2007) The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry*, **164**, 942–948.
2. Faraone,S.V., Sergeant,J., Gillberg,C. and Biederman,J. (2003) The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry*, **2**, 104–113.

3. Biederman,J. and Faraone,S.V. (2005) Attention-deficit hyperactivity disorder. *Lancet*, **366**, 237–248.

4. Biederman,J., Newcorn,J. and Sprich,S. (1991) Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *Am. J. Psychiatry*, **148**, 564–577.

5. Faraone,S.V., Perlis,R.H., Doyle,A.E., Smoller,J.W., Goralnick,J.J., Holmgren,M.A. and Sklar,P. (2005) Molecular genetics of attention-deficit/hyperactivity disorder. *Biol. Psychiatry*, **57**, 1313–1323.

6. Tripp,G. and Wickens,J.R. (2009) Neurobiology of ADHD. *Neuropharmacology*, **57**, 579–589.

7. Acosta,M.T., Arcos-Burgos,M. and Muenke,M. (2004) Attention deficit/hyperactivity disorder (ADHD): complex phenotype, simple genotype? *Genet. Med.*, **6**, 1–15.

8. Gizer,I.R., Ficks,C. and Waldman,I.D. (2009) Candidate gene studies of ADHD: a meta-analytic review. *Hum. Genet.*, **126**, 51–90.

9. Zhou,K., Dempfle,A., Arcos-Burgos,M., Bakker,S.C., Banaschewski,T., Biederman,J., Buitelaar,J., Castellanos,F.X., Doyle,A., Ebstein,R.P. *et al.* (2008) Meta-analysis of genome-wide linkage scans of attention deficit hyperactivity disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 1392–1398.

10. Neale,B.M., Medland,S.E., Ripke,S., Asherson,P., Franke,B., Lesch,K.P., Faraone,S.V., Nguyen,T.T., Schafer,H., Holmans,P. *et al.* (2010) Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry*, **49**, 884–897.

11. Smalley,S.L., Kustanovich,V., Minassian,S.L., Stone,J.L., Ogdie,M.N., McGough,J.J., McCracken,J.T., MacPhie,I.L., Francks,C., Fisher,S.E. *et al.* (2002) Genetic linkage of attention-deficit/hyperactivity disorder on chromosome 16p13, in a region implicated in autism. *Am. J. Hum. Genet.*, **71**, 959–963.

12. Ogdie,M.N., Fisher,S.E., Yang,M., Ishii,J., Francks,C., Loo,S.K., Cantor,R.M., McCracken,J.T., McGough,J.J., Smalley,S.L. *et al.* (2004) Attention deficit hyperactivity disorder: fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11. *Am. J. Hum. Genet.*, **75**, 661–668.

13. Poelmans,G., Pauls,D.L., Buitelaar,J.K. and Franke,B. (2011) Integrated genome-wide association study findings: identification of a neurodevelopmental network for attention deficit hyperactivity disorder. *Am. J. Psychiatry*, **168**, 365–377.

14. Elia,J., Gai,X., Xie,H.M., Perin,J.C., Geiger,E., Glessner,J.T., D'Arcy,M., deBerardinis,R., Frackelton,E., Kim,C. *et al.* (2010) Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol. Psychiatry*, **15**, 637–646.

15. Williams,N.M., Zaharieva,I., Martin,A., Langley,K., Mantripragada,K., Fossdal,R., Stefansson,H., Stefansson,K., Magnusson,P., Gudmundsson,O.O. *et al.* (2010) Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet*, **376**, 1401–1408.

16. Lesch,K.P., Selch,S., Renner,T.J., Jacob,C., Nguyen,T.T., Hahn,T., Romanos,M., Walitza,S., Shoichet,S., Dempfle,A. *et al.* (2011) Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. *Mol. Psychiatry*, **16**, 491–503.

17. Brookes,K., Xu,X., Chen,W., Zhou,K., Neale,B., Lowe,N., Anney,R., Franke,B., Gill,M., Ebstein,R. *et al.* (2006) The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Mol. Psychiatry*, **11**, 934–953.

18. Johansson,S., Halleland,H., Halmoy,A., Jacobsen,K.K., Landaas,E.T., Dramsdahl,M., Fasmer,O.B., Bergsholm,P., Lundervold,A.J., Gillberg,C. *et al.* (2008) Genetic analyses of dopamine related genes in adult ADHD patients suggest an association with the DRD5–microsatellite repeat, but not with DRD4 or SLC6A3 VNTRs. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 1470–1475.

19. Piletz,J.E., Zhang,X., Ranade,R. and Liu,C. (2011) Database of genetic studies of bipolar disorder. *Psychiatr. Genet.*, **21**, 57–68.

20. Allen,N.C., Bagade,S., McQueen,M.B., Ioannidis,J.P., Kavvoura,F.K., Khoury,M.J., Tanzi,R.E. and Bertram,L. (2008) Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat. Genet.*, **40**, 827–834.

21. Jia,P., Sun,J., Guo,A.Y. and Zhao,Z. (2010) SZGR: a comprehensive schizophrenia gene resource. *Mol. Psychiatry*, **15**, 453–462.

22. Basu,S.N., Kollu,R. and Banerjee-Basu,S. (2009) AutDB: a gene reference resource for autism research. *Nucleic Acids Res.*, **37**, D832–D836.

23. Matuszek,G. and Talebizadeh,Z. (2009) Autism Genetic Database (AGD): a comprehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. *BMC Med. Genet.*, **10**, 102.

24. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.

25. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2011) Ensembl 2011. *Nucleic Acids Res.*, **39**, D800–D806.

26. Seal,R.L., Gordon,S.M., Lush,M.J., Wright,M.W. and Bruford,E.A. (2011) genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.

27. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.

28. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

29. Kanehisa,M., Goto,S., Hattori,M., Aoki-Kinoshita,K.F., Itoh,M., Kawashima,S., Katayama,T., Araki,M. and Hirakawa,M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.

30. Lander,E. and Kruglyak,L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.*, **11**, 241–247.

31. Coghill,D. and Banaschewski,T. (2009) The genetics of attention-deficit/hyperactivity disorder. *Expert Rev. Neurother.*, **9**, 1547–1565.

32. Altshuler,D.M., Gibbs,R.A., Peltonen,L., Dermitzakis,E., Schaffner,S.F., Yu,F., Bonnen,P.E., de Bakker,P.I., Deloukas,P., Gabriel,S.B. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58.

33. Lasky-Su,J., Neale,B.M., Franke,B., Anney,R.J., Zhou,K., Maller,J.B., Vasquez,A.A., Chen,W., Asherson,P., Buitelaar,J. *et al.* (2008) Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 1345–1354.

34. Lesch,K.P., Timmesfeld,N., Renner,T.J., Halperin,R., Roser,C., Nguyen,T.T., Craig,D.W., Romanos,J., Heine,M., Meyer,J. *et al.* (2008) Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J. Neural Transm.*, **115**, 1573–1585.

35. Neale,B.M., Lasky-Su,J., Anney,R., Franke,B., Zhou,K., Maller,J.B., Vasquez,A.A., Asherson,P., Chen,W., Banaschewski,T. *et al.* (2008) Genome-wide association scan of attention deficit hyperactivity disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 1337–1344.

36. Mick,E., Todorov,A., Smalley,S., Hu,X., Loo,S., Todd,R.D., Biederman,J., Byrne,D., Dechairo,B., Guiney,A. *et al.* (2010) Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child. Adolesc. Psychiatry*, **49**, 898–905 e893.

37. Neale,B.M., Medland,S., Ripke,S., Anney,R.J., Asherson,P., Buitelaar,J., Franke,B., Gill,M., Kent,L., Holmans,P. *et al.* (2010) Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child. Adolesc. Psychiatry*, **49**, 906–920.

38. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment

analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, **102**, 15545–15550.

39. Zhang,K., Cui,S., Chang,S., Zhang,L. and Wang,J. (2010) i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.*, **38**, W90–W95.

40. Zhang,K., Chang,S., Cui,S., Guo,L., Zhang,L. and Wang,J. (2011) ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res.*, **39**, W437–W443.

41. Mailman,M.D., Feolo,M., Jin,Y., Kimura,M., Tryka,K., Bagoutdinov,R., Hao,L., Kiang,A., Paschall,J., Phan,L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.

42. Liberzon,A., Subramanian,A., Pinchback,R., Thorvaldsdottir,H., Tamayo,P. and Mesirov,J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.

43. Ferreira,M.A., O'Donovan,M.C., Meng,Y.A., Jones,I.R., Ruderfer,D.M., Jones,L., Fan,J., Kirov,G., Perlis,R.H., Green,E.K. *et al.* (2008) Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.*, **40**, 1056–1058.

44. Askland,K., Read,C. and Moore,J. (2009) Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum. Genet.*, **125**, 63–79.

45. Pal,P., Mihanovic,M., Molnar,S., Xi,H., Sun,G., Guha,S., Jeran,N., Tomljenovic,A., Malnar,A., Missoni,S. *et al.* (2009) Association of tagging single nucleotide polymorphisms on 8 candidate genes in dopaminergic pathway with schizophrenia in Croatian population. *Croat. Med. J.*, **50**, 361–369.

46. Cubells,J.F., Sun,X., Li,W., Bonsall,R.W., McGrath,J.A., Avramopoulos,D., Lasseter,V.K., Wolyniec,P.S., Tang,Y.L., Mercer,K. *et al.* (2011) Linkage analysis of plasma dopamine beta-hydroxylase activity in families of patients with schizophrenia. *Hum. Genet*, **130**, 635–643.

47. Punia,S., Das,M., Behari,M., Mishra,B.K., Sahani,A.K., Govindappa,S.T., Jayaram,S., Muthane,U.B., K,T.B. and Juyal,R.C. (2010) Role of polymorphisms in dopamine synthesis and metabolism genes and association of DBH haplotypes with Parkinson's disease among North Indians. *Pharmacogenet. Genomics*, **20**, 435–441.

48. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

49. Stergiakouli,E. and Thapar,A. (2010) Fitting the pieces together: current research on the genetic basis of attention-deficit/hyperactivity disorder (ADHD). *Neuropsychiatr. Dis. Treat*, **6**, 551–560.

50. Wang,K., Li,M. and Hakonarson,H. (2010) Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.*, **11**, 843–854.